Solution structure of a GAAA tetraloop receptor RNA

Samuel E.Butcher, Thorsten Dieckmann and Juli Feigon[1]

Department of Chemistry and Biochemistry, and Molecular Biology Institute, University of California, Los Angeles, CA 90095-1569, USA

[1]Corresponding author

The GAAA tetraloop receptor is an 11-nucleotide RNA sequence that participates in the tertiary folding of a variety of large catalytic RNAs by providing a specific binding site for GAAA tetraloops. Here we report the solution structure of the isolated tetraloop receptor as solved by multidimensional, heteronuclear magnetic resonance spectroscopy. The internal loop of the tetraloop receptor has three adenosines stacked in a cross-strand or zipper-like fashion. This arrangement produces a high degree of base stacking within the asymmetric internal loop without extrahelical bases or kinking the helix. Additional interactions within the internal loop include a U·U mismatch pair and a G·U wobble pair. A comparison with the crystal structure of the receptor RNA bound to its tetraloop shows that a conformational change has to occur upon tetraloop binding, which is in good agreement with previous biochemical data. A model for an alternative binding site within the receptor is proposed based on the NMR structure, phylogenetic data and previous crystallographic structures of tetraloop interactions.
Keywords: GAAA tetraloop/group I intron/NMR/ribozyme/RNA structure/tetraloop receptor

Introduction

Long-range tertiary interactions are essential for the proper folding and function of large, biologically active RNAs. A highly conserved 11-nucleotide motif containing an internal loop, termed the tetraloop receptor, is known to mediate RNA tertiary folding by providing a binding site for GAAA tetraloops (Costa and Michel, 1995). The GAAA tetraloop receptor domain has been identified in group I and II introns, as well as in the RNase P of some Gram-positive bacteria (Tanner and Cech, 1995). The conservation of the tetraloop receptor domain throughout the evolution of these RNAs suggests an essential role for most of the 11 nucleotides within the motif. *In vitro* selection experiments using randomized tetraloop receptors revealed that the majority of cloned sequences harbored the canonical motif, with some minor sequence variants that include a C to A substitution within the loop and the exchange of a G·U for an A·C wobble pair (Costa and Michel, 1997). A nearly identical pattern of sequence conservation is found in nature among the group I and II introns and RNase P RNAs that contain the receptor domain (Tanner, 1997). The structure of the tetraloop receptor bound to its cognate GAAA tetraloop, as seen in the crystal structure of the P4–P6 domain of the *Tetrahymena* group I ribozyme, reveals that most of the conserved nucleotides within the receptor are involved in forming a specific interface with the tetraloop, stabilized by both stacking and hydrogen bonding (Cate *et al.*, 1996a). The three adenosines in the GAAA tetraloop were observed to bind to the tetraloop receptor via a base triplet, a quadruplet, base–sugar and sugar–sugar interactions. Within the receptor, an unusual structural motif was observed, comprised of two consecutive adenines in a co-planar or 'platform' arrangement upon which the GAAA tetraloop stacks. The adenosine platform motif was also observed at sites of intermolecular interactions within the crystal lattice, suggesting that adenosine platforms are a general motif for the mediation of RNA tertiary interactions (Cate *et al.*, 1996b).

Most, but not all, of the phylogenetic conservation of the tetraloop receptor sequence can be explained by the interactions observed in the P4–P6 crystal structure (Cate *et al.*, 1996a). For example, the C to A variant in the loop mentioned above is capable of forming an AC platform, a structure that is nearly identical to the AA platform (Zimmermann *et al.*, 1997). The only conserved sequence element which has yet to be structurally rationalized is the C·G pair at the first position (positions 3·21 in our numbering scheme or 222·251 in the *Tetrahymena* group I intron). Among 35 clones obtained from *in vitro* selection experiments, 28 had a C·G pair at this position (Costa and Michel, 1997), though no interactions with these bases were found in the P4–P6 crystal structure (Cate *et al.*, 1996a).

The basis for macromolecular discrimination between RNA and proteins has been elucidated for a number of systems by comparing the free and bound conformations of the components. In many cases, the interactions between proteins or peptides and RNA result in a conformational rearrangement of the RNA. For example, interactions involving tRNA[Asp] anticodon-synthetase (Ruff *et al.*, 1991), the U1A complex (Allain *et al.*, 1996), the HIV RRE–REV (Battiste *et al.*, 1996; Peterson and Feigon, 1996) and TAT–TAR (Aboul-ela *et al.*, 1995) all involve induced fit or conformational rearrangement upon binding. Similar results are observed for RNA aptamer complexes bound to their ligands (Dieckmann *et al.*, 1996; Jiang *et al.*, 1996; Zimmermann *et al.*, 1997). It remains to be seen whether RNA–RNA tertiary interactions also generally proceed through conformational rearrangements.

We have used NMR methods to solve the solution structure of the isolated tetraloop receptor domain, which consists of the phylogenetically conserved 11-nucleotide consensus sequence (Costa and Michel, 1995) embedded within a model 23-nucleotide stem–loop sequence. Instead
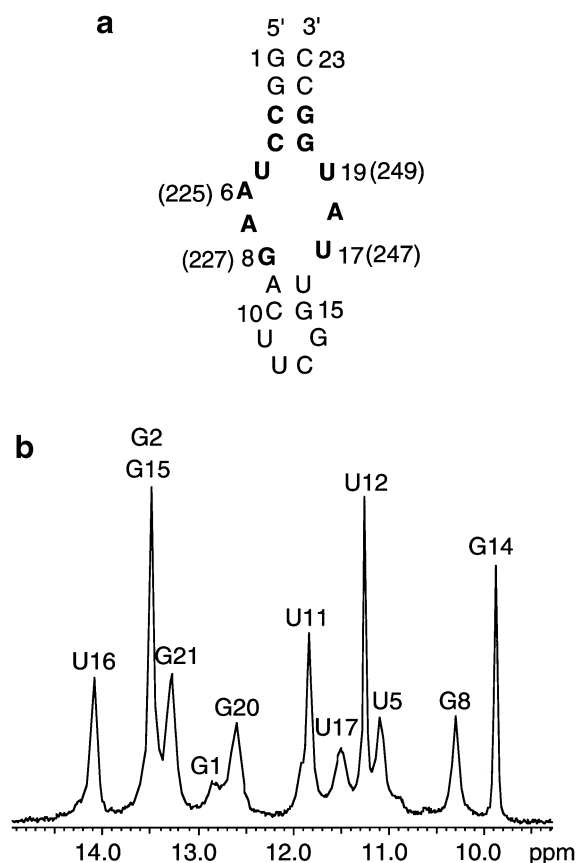
Fig. 1. (a) Sequence and secondary structure of the tetraloop receptor RNA used in this study. The 11 conserved nucleotides for the receptor are in bold. The numbering system is indicated, with the corresponding numbering scheme for the *Tetrahymena* group I intron tetraloop receptor in parenthesis. (b) 500 MHz 1D ^{1}H NMR spectrum of the imino region of the tetraloop receptor RNA recorded at 274 K. Ninety-six scans of 4096 points were acquired, with a sweep width of 10 000 Hz. The data were zero filled to 8192 points and processed with an exponential filter function with a line broadening of 3.0 Hz. The sample was 1 mM in 450 μl, pH 5.5, in 90% H_2O/10% D_2O and 100 mM NaCl. Assignments of the imino resonances are indicated.

of the A platform found in the crystal structure of the bound form of the tetraloop receptor (Cate *et al.*, 1996a), the adenines are arranged in a cross-strand stacking arrangement that we call a base zipper. We use molecular modeling to show that an alternative tetraloop-binding site exists within the receptor, based upon the phylogenetic conservation of the tetraloop receptor sequence and previously determined crystallographic structures of GAAA tetraloop interactions (Pley *et al.*, 1994a; Cate *et al.*, 1996a).

Results

Assignment of proton resonances
The sequence of the GAAA tetraloop receptor RNA is shown in Figure 1a. The numbering scheme used in this study is indicated, while the numbering system for the *Tetrahymena* group I intron is shown in parenthesis (Burke *et al.*, 1987). The two strands were linked by the extra stable UUCG tetraloop, which has been well characterized by NMR (Cheong *et al.*, 1990; Allain and Varani, 1995b). Titrations of up to 25 mM $MgCl_2$ indicated no evidence for magnesium-induced conformational changes; therefore,

most spectra were acquired in the presence of 100 mM NaCl. The base-paired imino proton resonances in the stem regions and the UUCG tetraloop (Cheong *et al.*, 1990) are clearly visible, as are the U5, G8 and U17 imino proton resonances (Figure 1b). The exchangeable protons were assigned via sequential NOEs observed in a 2D NOESY spectrum, and were confirmed with ^{1}H-^{15}N HMQC, ^{1}H-^{15}N 2D HMQC-NOESY and 2D HCCCNH TOCSY spectra as previously described (Dieckmann and Feigon, 1997) (data not shown). In the internal loop, a strong imino–imino NOE is observed for G8 and U17 which, in combination with the non-exchangeable NOE data, is consistent with the formation of a G·U wobble pair closing the internal loop. The U5 imino resonance in the internal loop is partially protected from solvent exchange, while the U19 imino resonance is not visible and is therefore exposed to solvent.

Complete assignments for all of the non-exchangeable proton resonances and their directly bound carbons were obtained using a series of experiments in D_2O including homonuclear 2D NOESY, DQF COSY and TOCSY, as well as heteronuclear HCNCH, HCCH TOCSY optimized for adenine H8–H2 correlation, ^{1}H-^{13}C HSQC, ^{15}N long-range HSQC, 3D (^{1}H-^{13}C) HMQC-NOESY and 3D HCCH TOCSY and COSY experiments following previously described protocols (Nikonowicz and Pardi, 1993; Dieckmann and Feigon, 1994, 1997; Pardi, 1995; Varani *et al.*, 1996) (see Materials and methods). Several starting points were available for assigning the two stem regions by their sequential NOE connectivities. Both the UUCG tetraloop and the terminal stem sequence beginning with consecutive G·C base pairs displayed nearly identical chemical shifts to those previously reported (Allain and Varani, 1995a). While the single adenine H2 resonance in the stem (A9) is easily assignable by its sharp, characteristic NOE to the U16 imino proton, the three internal loop adenine H2 protons could only be unambiguously assigned by direct correlation with their corresponding H8 protons, using an HCCH TOCSY experiment optimized for adenine H8–H2 correlation (Legault *et al.*, 1994; Marino *et al.*, 1994). Figure 2 shows a ^{1}H-^{13}C HSQC aligned with the HCCH TOCSY, demonstrating the through-bond correlation between the adenine H2 and H8 resonances. Interestingly, two of the adenine H2 protons in the internal loop (A6 and A18) resonate upfield of the stem A9 H2 proton. These unusually high field chemical shifts may be caused by ring current effects, if the A H2 protons are stacked directly above or below an aromatic ring.

Sequential NOE connectivities can be traced from the stem regions throughout the internal loop without interruption (Figure 3). The sequential sugar-to-base and base-to-base internucleotide NOEs observed in the 2D NOESY spectrum indicate that there are no extrahelical bases in the internal loop. Analysis of the NOE data as a function of NOESY mixing time suggest that the three spacings between the bases of A6 and A7, U17 and A18, and A18 and U19 are larger than those found in A-form RNA, since the aromatic internucleotide NOEs for these bases are less intense than the ones observed for bases within the Watson–Crick stems.

Torsion angles
Analysis of the short mixing time NOESY spectra (50 ms) indicates that all of the bases in the internal loop have
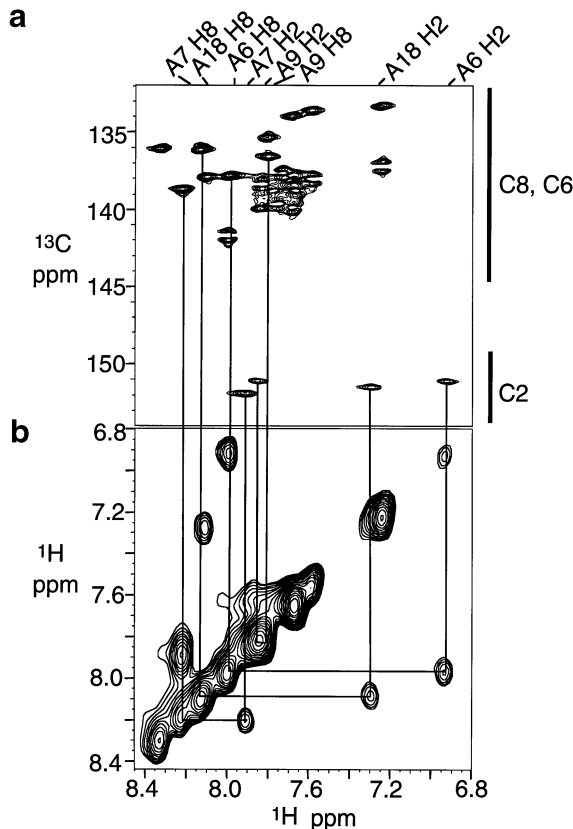
a

b

Fig. 2. (**a**) Aromatic portion of a 600 MHz ^{1}H-^{13}C HSQC spectrum. A total of 1024 and 512 points were acquired in t_2 and t_1, respectively, with 16 scans per t_1 increment, and a relaxation delay of 1.6 s. The sweep width was 6000 Hz in t_2 and 12 000 Hz in t_1. The final data matrix was 1024×1024 points and was processed with a 90° shifted squared sine bell filter function. (**b**) 500 MHz ^{1}H-^{13}C HCCH TOCSY spectrum optimized for AH2–AH8 correlations. 512 and 160 points were acquired in t_2 and t_1, respectively, with 320 scans per t_1 increment, and a relaxation delay of 1.6 s. The sweep width was 5000 Hz in both dimensions. The final data matrix was 1024×1024 points and was processed with a 90° shifted squared sine bell filter function. Lines trace the through-bond correlations between AH2 and AH8 protons. Spectra were recorded at 293 K. The fully ^{13}C,^{15}N-labeled RNA was 1 mM in 200 μl in a Shigemi NMR tube, pH 6.2, 100 mM NaCl in D$_2$O.

Fig. 3. H1′ to aromatic region of a NOESY spectrum (τ_m = 250 ms) of the tetraloop receptor RNA at 303 K in D$_2$O showing the H1′-base H8/H6 crosspeak region. The base-H1′ sequential connectivities are traced. A total of 1024 and 800 complex points were acquired in t_2 and t_1, respectively, with 96 scans per t_1 increment, and a relaxation delay of 1.6 s. The sweep width was 5000 Hz in both dimensions. The final data matrix was 2048×2048 points and was processed with a Gaussian filter function (line broadening –18 Hz, GB 0.08 in f$_2$ and 0.14 in f$_1$). RNA sample is the same as in Figure 1, except that the pH was raised to 6.2 and the sample transferred into D$_2$O.

Table I. Structure determination statistics for the 20 lowest energy tetraloop receptor RNA structures

NOE distance restraints	318
Internucleotide NOEs	182
Intranucleotide NOEs	136
Average NOEs per core residue (3–8,17–21)	19.6
R.m.s.d. for all core residue heavy atoms relative to the mean structure (Å)	1.03 ± 0.28
R.m.s.d. for all heavy atoms relative to the mean structure (Å)	2.26 ± 0.59
NOE violations (Å)	0 >0.2
Angle violations (°)	0 >5
Mean deviation from covalent geometry	
Bond lengths (Å)	0.01
Angles (°)	2.6
Impropers (°)	0.79

glycosidic angles in the *anti* range, with weak intranucleotide base to H1′ NOEs and strong intranucleotide base to H3′ NOEs (data not shown). The single exception in the molecule is the *syn* base G14, which is in the UUCG tetraloop (Allain and Varani, 1995b; also data not shown). In addition, the weak or absent H1′–H2′ crosspeaks observed in DQF COSY spectra (J$_{H1',H2}$ <2 Hz) indicate that most of the sugar puckers are C3′-endo, with the exception of the two sugars in the UUCG tetraloop known to be C2′-endo (Allain and Varani, 1995b). Larger (4–8 Hz) H1′–H2′ couplings inconsistent with a pure C3′-endo sugar conformation were also observed for three of the internal loop nucleotides (A6, U17 and A18), as well as the terminal G1 and C23 nucleotides. Of these nucleotides, A18 has the largest (8 Hz) H1′–H2′ coupling, suggesting that this sugar pucker may in fact be predominantly S-type. The sugar puckers for A6, A7, G8, U17 and A18 were left unrestrained in the structure calculations so as not to bias the outcomes of the sugar conformations. Additional information about sugar conformation is
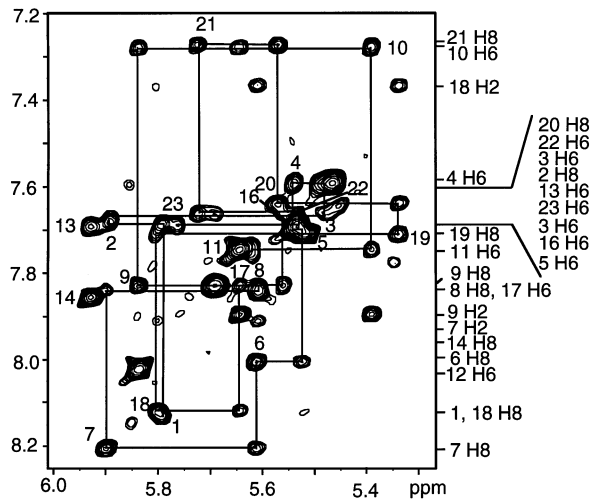
obtained from the C1′ chemical shifts, observed in the ^{1}H-^{13}C HSQC spectrum. The G1, A6, A7 and A18 C1′ chemical shifts, as well as those belonging to the two C2′-endo sugars in the UUCG tetraloop, were all within the S-type (C2′-endo) range (88–91 p.p.m.), while the rest of the C1′ chemical shifts were within the N-type (C3′-endo) sugar range (91–94 p.p.m.). Test calculations in which A6, A7 and A18 were restrained to be C2′-endo were found to completely satisfy the NMR data (data not shown).

Structure of the GAAA tetraloop receptor

A total of 318 NOE-derived distance constraints were obtained. The average number of NOE distance restraints within the internal loop region is 19.6 NOEs per nucleotide (Table I). The large number of distance constraints obtained for these residues helps to define precisely the orientation of the bases in the internal loop. In total, 100 starting structures were calculated by distance geometry and
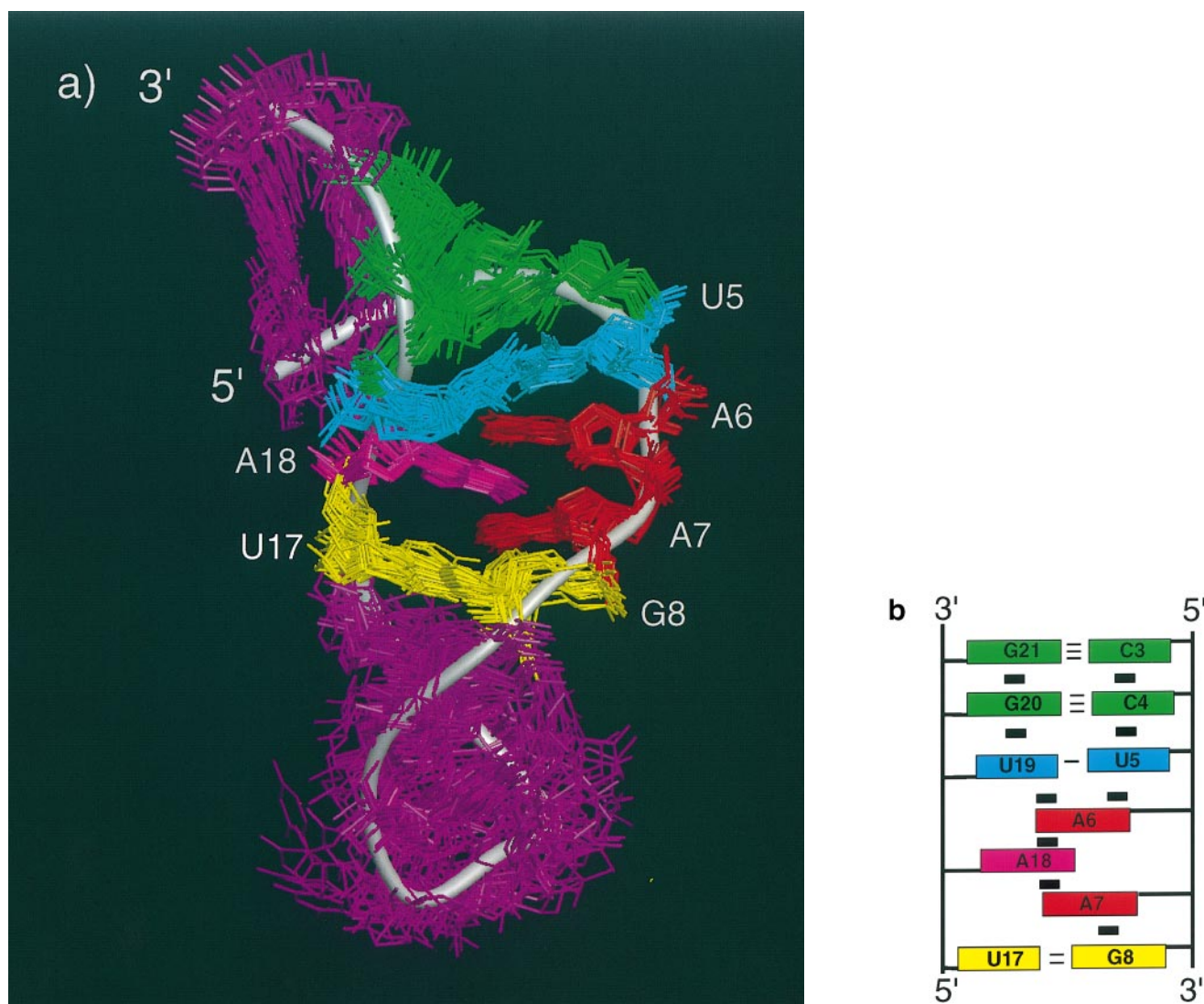
Fig. 4. (**a**) The 20 lowest energy structures, superimposed upon the heavy atoms of nucleotides 3–8 and 17–21. View is into the minor groove. Non-conserved stem nucleotides and the UUCG tetraloop are purple. The conserved tandem C·G pairs (nucleotides 3–4 and 20–21) are green, U5 and U19 are cyan, A6 and A7 are red, the G8·U17 wobble is yellow and A18 is magenta. (**b**) Schematic illustration of the tetraloop receptor RNA. Rectangles indicate bases, and stacking interactions are shown as filled black rectangles. Hydrogen bonds are indicated with black lines.

simulated annealing methods using X-PLOR (Brünger, 1992), as described in Materials and methods. The distance geometry structures were refined using restrained molecular dynamics and simulated annealing. The 20 lowest energy structures out of the 100 calculated were subjected to a final round of molecular dynamics refinement and were evaluated.

The 20 lowest energy structures are shown in Figure 4a. The conserved 11-nucleotide receptor region is well defined in all 20 low-energy structures and has an r.m.s.d. value for all heavy atoms relative to the mean structure of 1.03 ± 0.28 Å (Table I). At the bottom of the receptor internal loop, a G·U wobble pair forms. The internal loop adenines are arranged in an unusual cross-strand stacking or 'base zipper' motif, with A18 reaching across the internal loop and stacking between A6 and A7, while A6 stacks between A18 and U5 (Figure 4a and b). No hydrogen bonds occur within this zipper region, which appears to be stabilized only by the stacking interactions. The cross-strand stacking interactions in the internal loop

are between the six-membered rings of the adenines. The uridines at the top of the internal loop stack between the stem C·G pair and A6, and form a base pair in the majority of the structures (12 out of 20) consisting of a single hydrogen bond between the imino proton of U5 and the O4 oxygen of U19.

The structure correlates very well with NMR data that were not included in the calculations, such as proton exchange rates and chemical shifts. The U5 imino proton is hydrogen-bonded to the O4 carbonyl oxygen of U19 in the majority of the structures, and this imino proton is observed to exchange slowly with the solvent. Conversely, the U19 imino proton points directly out to the solvent, and this imino proton exchanges too rapidly to be observed in 1D NMR spectra. The cross-strand stacking arrangement of the zipper motif provides an explanation for the observed chemical shifts of the AH2 protons. The upfield shifted A6 and A18 H2 protons are stacked directly upon the six-membered rings of A18 and A7, respectively, where they would be subject to ring current effects, while
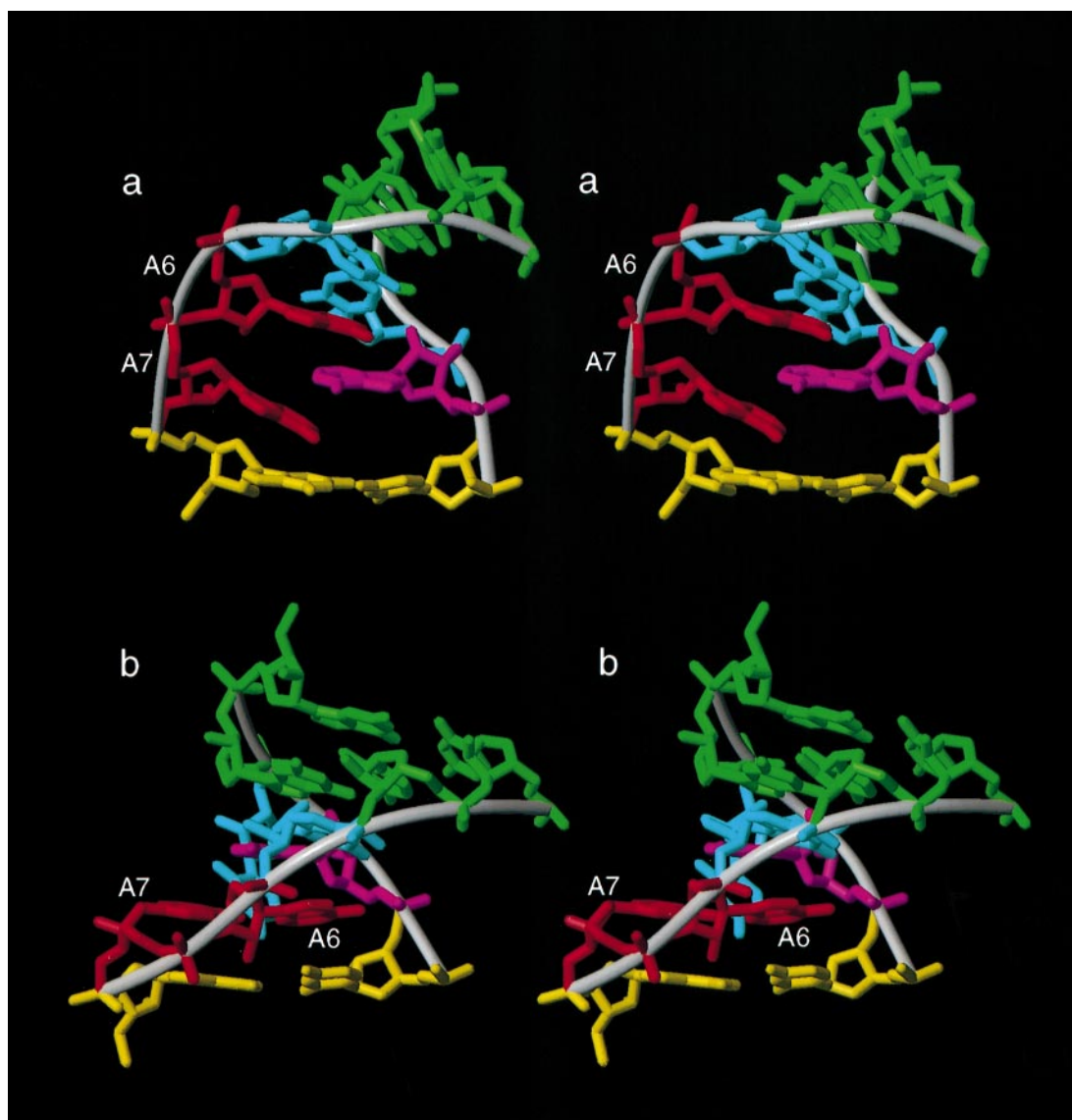
Fig. 5. Comparison of the structures of the (a) free and (b) bound forms of the tetraloop receptor: (**a**) is the lowest energy NMR structure of the free receptor; (**b**) is the crystal structure of the bound tetraloop receptor as observed in the P4–P6 domain of the *Tetrahymena* group I ribozyme (Cate *et al.*, 1996a). For clarity, the tetraloop is not shown. View is into the major groove.

the downfield shifted A7 H2 is exposed to solvent and not stacked. Finally, the chemical shifts observed in [1]H-^{15}N HMQC spectra indicate that the adenine amino groups in the internal loop are not hydrogen-bonded, and a ^{1}H-^{15}N HSQC (Sklenář *et al.*, 1994) spectrum suggests that there are no hydrogen-bonded N7 atoms (data not shown).

Discussion

Comparison of the structures of the free and bound GAAA tetraloop receptor RNA

The isolated GAAA tetraloop receptor forms a structure that is quite different from that of the bound tetraloop receptor, which folds into an adenosine platform motif (Cate *et al.*, 1996a). A comparison of the two structures is shown in Figure 5a and b. There are some structural elements common to both structures, including the G·U wobble pair at the bottom and the tandem C·G pairs at the top of the internal loop. Additionally, A7, which

corresponds to the 3′ A in the platform motif, stacks upon the G in the G·U wobble pair in both structures, and U5 stacks upon C4 in both structures.

Major differences between the two structures include the location of the three adenines (A6, 7 and 18) and U19. In the free tetraloop receptor, A6 stacks between U5 and A18, while in the bound form of the receptor, A6 would stack upon U17 to form a portion of the adenosine platform. In the solution structure of the free receptor, U19 stacks upon the stem G20, while in the bound form A18 would stack upon G20 and U19 would be extrahelical. The many cross-strand NOEs that we observe in the internal loop precisely define the base zipper and are clearly inconsistent with the formation of an adenosine platform as observed in the P4–P6 crystal structure (Cate *et al.*, 1996a).

The solution structure of the free GAAA tetraloop receptor is in good agreement with biochemical data which suggest that a conformational change takes place upon

removing the cognate tetraloop (Murphy and Cech, 1994; Cate *et al.*, 1996a). Murphy and Cech demonstrated that mutations in the P4–P6 domain that disrupt the receptor–tetraloop interaction give rise to increased dimethylsulfate (DMS) reactivity at the adenine N1 at position 225 (corresponding to A6 in our numbering scheme), while in the bound form of the receptor this position is protected from DMS modification. Consistent with this data, we find that the A6 N1 position is accessible to solvent in the solution structure of the free receptor, which correlates nicely with the enhanced DMS reactivity at this position when the receptor is probed in its unbound form. Conversely, the A7 N1 (corresponding to the 3′ A226 in the platform), which shows no enhanced reactivity to DMS in the free tetraloop receptor (Murphy and Cech, 1994), is buried in the solution structure.

The base zipper is a common structural motif

Interstrand stacking interactions are a common structural element in nucleic acids. Several examples of cross-strand stacking interactions exist for both RNA (Szewczak *et al.*, 1993; Wimberly *et al.*, 1993; Pley *et al.*, 1994b; Scott *et al.*, 1995) and DNA (Maskos *et al.*, 1993; Chou *et al.*, 1997). However, the RNA examples are of single cross-strand stacks and not multiple or 'zippered' stacks. Similar base zipper or 'interdigitation' motifs were observed over 20 years ago in tRNA[Phe], where interstrand stacking occurs for the six-membered ring of A9 between the six-membered rings of G45 and 46, as well as A21 between G46 and C48, and G18 between G57 and A58 (Robertus *et al.*, 1974). In these base zipper motifs, the 3′ nucleotide on one strand commonly adopts a C2′-endo conformation to allow the accommodation of an interdigited base between sequential nucleotides (Saenger, 1984). We note that spectroscopic data on A18 indicates that it probably is in the S-type conformational range most of the time, which is consistent with the fact that the backbone has to traverse a greater distance at this position to accommodate A7 between U17 and A18. A7 also has a C1′ chemical shift consistent with an S-type sugar pucker, although a range of sugar puckers at this position were obtained from the calculations. Concurrent with this work, another base zipper has been recently identified in the structure of a theophylline-binding RNA aptamer (Zimmermann *et al.*, 1997). Thus, it seems that base zippers are a commonly used RNA structural motif, and it will be interesting to determine how frequently they occur in other RNA internal loop structures.

How does a base zipper convert to a platform?

The fact that the structure of the free receptor is different from its structure in the context of the tertiary interactions in the P4–P6 crystal structure argues for a conformational rearrangement of the tetraloop receptor upon binding the GAAA tetraloop. Initial folding of the receptor into the base zipper motif appears likely, if one assumes that helical elements fold more quickly than long-range tertiary interactions. In the case of the RNase P tetraloop receptor (Tanner and Cech, 1995), the receptor must fold before the tetraloop, because the receptor portion of the RNA is 5′ to the cognate tetraloop and is transcribed first. The conformational rearrangement required to form the adenosine platform motif is illustrated in Figure 5. The backbone of one strand of the molecule must slide across the major groove to unstack the zipper motif and form the co-planar consecutive adenosines that make up the platform, upon which the tetraloop stacks. During adenosine platform formation, the single hydrogen bond between the U·U mismatch pair must be broken and the 3′U bulged out of the helix.

The sequence conservation of the top C3·G21 base pair of the tetraloop receptor has yet to be explained. In the crystal structure, only a single 2′OH from these nucleotides is within hydrogen bonding distance of the bound tetraloop (Cate *et al.*, 1996a). Therefore, the crystal structure does not explain the high degree of conservation of this base pair in nature, which argues for an important function associated with this base pair. Simple stability arguments do not provide a sufficient explanation, since nearest neighbor rules indicate that a C·G pair at this position is less stable than G·C (Serra and Turner, 1995).

Interestingly, it has been known for some time that GAAA tetraloops co-vary and interact with tandem C·G pairs (Michel and Westhof, 1990; Jaeger *et al.*, 1994). This interaction has been observed directly in a hammerhead ribozyme crystal structure, in which an interaction between a GAAA tetraloop and tandem C·G pairs was observed within the minor groove of an RNA helix (Pley *et al.*, 1994a). The interaction between the tetraloop and the tandem stem C·G pairs involves a (G·A)·(C·G) base quadruple with the first C·G pair (Pley *et al.*, 1994a). This base quadruple interaction has exactly the same hydrogen bond interactions as the one observed in the P4–P6 crystal structure (Cate *et al.*, 1996a), except that the P4–P6 interactions are with the second C·G pair instead of the first. In other words, the docking of the tetraloops into the tandem C·G pairs differs in register by one C·G base pair between the two crystal structures. Therefore, it is possible that the phylogenetic conservation of the first C·G pair in the tetraloop receptor is to maintain an alternative docking register for the tetraloop.

The presence of an alternative docking site for the tetraloop suggests that the interaction with the receptor may not require the formation of an adenosine platform *a priori*. We propose a model in which the tetraloop initially recognizes the receptor by docking at the first C·G base pair and forming the base quadruplet observed in the hammerhead crystal structure (Pley *et al.*, 1994a). Such an interaction would also be stabilized by a base triple between the second C·G pair and the first A in the GAAA tetraloop as described (Pley *et al.*, 1994a). This could nucleate a structural transition in which the GAAA tetraloop translocates down to the second C·G base pair to form the same type of base quadruplet, and the receptor conformation converts from the A zipper to the A platform. Molecular modeling calculations show that binding of a GAAA tetraloop in the alternative register is sterically feasible. The proposed initial interaction is shown (Figure 6a and c). The model was generated by energy minimization using X-PLOR and the hydrogen bonds described by Pley *et al.* (1994a) as distance constraints. The position of the tetraloop in the model differs from the one in the crystal structure of Cate *et al.* (1996a) by one nucleotide in register (Figure 6b and d). Aside from assigning a role for an apparently unrecognized though phylogenetically conserved C·G base pair, the model makes testable predic-
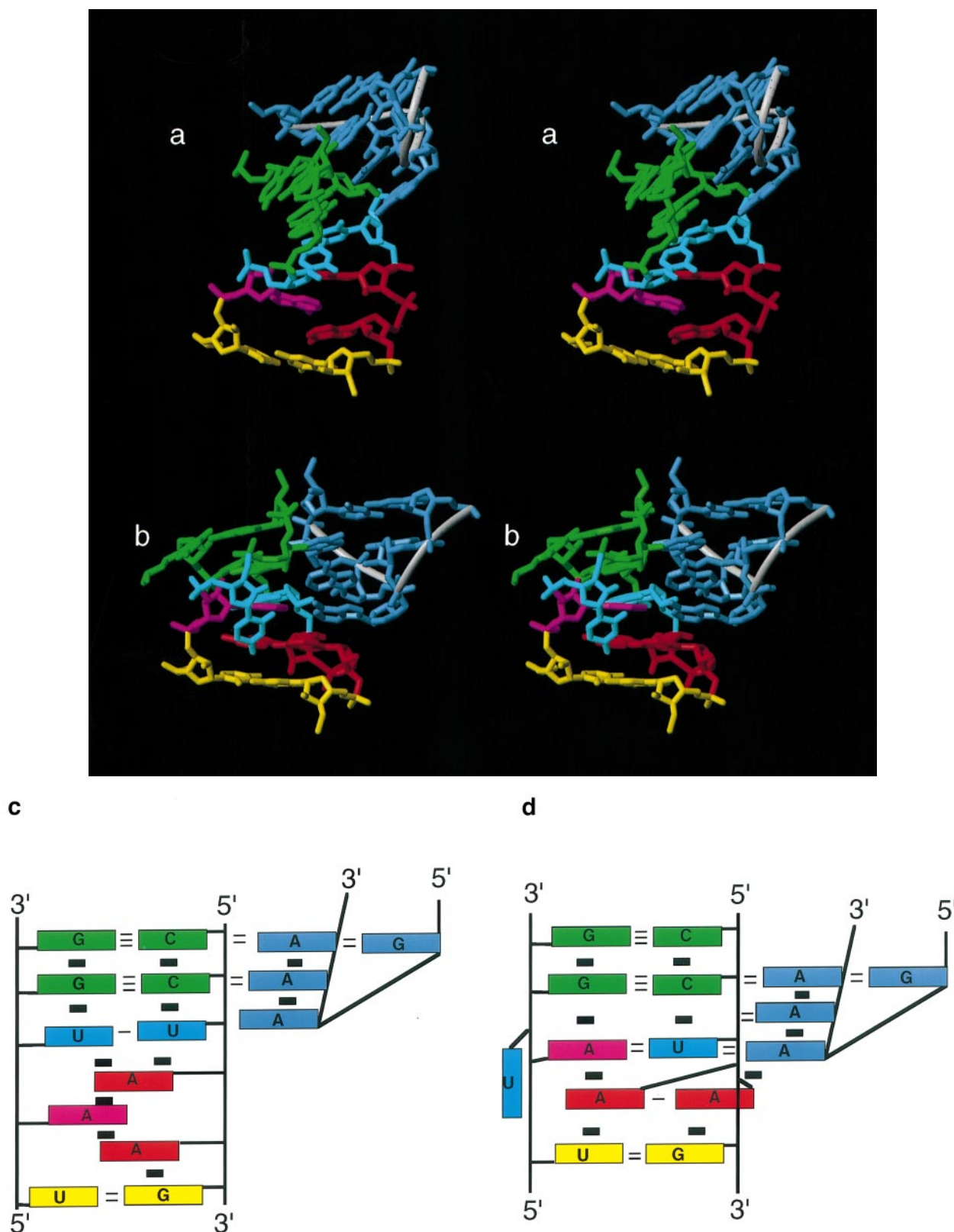
Fig. 6. Model for an alternative tetraloop-binding site within the receptor RNA. (**a**) The model of a GAAA tetraloop docked into the conserved C·G pairs of the receptor, created with a set of distance restraints based on the NMR structure and a previously observed tetraloop interaction (Pley *et al.*, 1994a). View is into the minor groove. (**b**) The crystal structure of the GAAA·receptor complex within the P4–P6 domain of the *Tetrahymena* group I ribozyme (Cate *et al.*, 1996a), shown as a comparison. (**c**) Schematic diagram of the interaction shown in (a). (**d**) Schematic diagram of the interaction shown in (b).

tions. If the alternative binding site is utilized during the course of folding or catalysis, then mutations at the top C·G pair may produce a measurable effect upon one of these steps, even though they would not be expected to interfere with adenosine platform formation. For example, the presence of the top C·G pair may increase the on-rate of the tetraloop for its receptor, if the conformational change is a rate-limiting step for tetraloop binding; or the alternative binding site may be utilized as part of a conformational switch required for a particular step of catalysis.

Materials and methods

Sample preparation

RNA was prepared enzymatically from a DNA template using T7 RNA polymerase (Milligan *et al.*, 1987) and unlabeled NTPs or ^{13}C, ^{15}N-labeled NTPs. Labeled NTPs were isolated from *Methylobacterium extorquens* bacteria strain AM1 which had been grown in media containing [^{13}C]methanol and [^{15}N]ammonia as the sole carbon and nitrogen sources, purified, and converted to NTPs as described (Batey *et al.*, 1992; Nikonowicz *et al.*, 1992; Peterson *et al.*, 1994). After transcription, Mg^{2+}-pyrophosphate was removed by brief centrifugation, and the RNA was ethanol-precipitated. RNA was gel purified on 15% polyacrylamide–8 M urea gels. The correct RNA band was identified by UV shadowing, excised from the gel and electroeluted. RNA was further purified by DEAE–Sepharose anion exchange chromatography followed by Sepharose G-15 gel filtration and lyophilized to dryness. NMR samples were 0.9 mM in RNA strand, 100 mM NaCl, pH 5.5 or 6.2 in 450 µl 90% H_2O/10% D_2O or D_2O (99.99% D_2O from Isotec). Samples were exchanged from 90% H_2O/10% D_2O to D_2O by drying under nitrogen in the NMR tube and resuspending in 99.99% D_2O.

NMR spectroscopy

Spectra were acquired at 500 and 600 MHz on Bruker DRX spectrometers. Solvent suppression for samples in 90% H_2O/10% D_2O was achieved using 1$\bar{1}$ spin echo pulse sequences (Sklenář *et al.*, 1987). The residual HDO resonance in D_2O samples was suppressed using low-power presaturation. Quadrature detection for the indirect dimensions of multidimensional experiments was achieved using the States-TPPI method (Marion *et al.*, 1989). Two-dimensional NOESY spectra (Macura *et al.*, 1980) in 90% H_2O/10% D_2O were acquired at 274 and 278 K with a mixing time of 150 ms. A 2D CITY-TOCSY (Bax and Davis, 1985; Kadkhodaei *et al.*, 1993) with a mixing time of 60 ms, a DQF-COSY (Rance *et al.*, 1983) and NOESY spectra with mixing times of 50, 100, 200, 250 and 300 ms in D_2O were measured at 293 K. 300 ms mixing time NOESY spectra were also acquired at 303 and 313 K.

Heteronuclear NMR spectra were measured at 293 K in 100% D_2O, with the exception of a ^{1}H-^{15}N HMQC which was acquired at 278 K in 90% H_2O/10% D_2O. ^{13}C and/or ^{15}N decoupling during the acquisition time was achieved using the GARP composite pulse sequence (Shaka *et al.*, 1985). Two-dimensional spectra acquired included a long-range ^{1}H-^{15}N HSQC (Sklenář *et al.*, 1994), ^{1}H-^{15}N HMQC, ^{1}H-^{13}C-constant time HSQC (Santoro and King, 1992), HCCH-TOCSY optimized for AH2–AH8 correlation (Legault *et al.*, 1994; Marino *et al.*, 1994), HCNCH (Sklenář *et al.*, 1993) and a ^{1}H-^{31}P-HETCOR (Sklenář *et al.*, 1986). Three-dimensional spectra acquired included 3D-HCCH-COSY (Clore *et al.*, 1990), 3D-HCCH-TOCSY (Bax *et al.*, 1990), and a 3D-NOESY-^{1}H-^{13}C-HMQC (Marion *et al.*, 1989; Nikonowicz and Pardi, 1993) with a mixing time of 300 ms. Additional acquisition and processing parameters are given in the figure legends. Data processing and analysis were performed using the software packages XWINNMR 1.2 and AURELIA 2.0 (Bruker Inc., Rheinstetten, Germany).

Magnesium titrations were performed to ascertain whether any evidence for magnesium-dependent structure could be obtained. ^{1}H-^{13}C HSQC spectra were recorded at various magnesium concentrations up to 25 mM and no changes in proton or carbon chemical shifts were observed upon the addition of magnesium. Therefore, most spectra were acquired in 100 mM NaCl, since a UpA step occurs within the internal loop, and UpA steps are known to be 'hot-spots' for magnesium hydrolysis (Puglisi and Wyatt, 1995). ^{1}H-^{13}C HSQC spectra were also recorded with varying pH values of 5.2, 5.5 and 6.2, and no evidence for protonation effects was observed (Legault and Pardi, 1994).

Input restraints for structure calculations

Exchangeable NOE distances were obtained from a 2D NOESY in 90% H_2O/10% D_2O at 274 K with a 150 ms mixing time. NOE distances for exchangeable protons were semi-quantitatively classified as strong, medium or weak. For these NOEs, conservative upper bounds were used corresponding to 3.5 Å (strong), 5 Å (medium) and 6 Å (weak), with the lower bounds equal to the sum of the van der Waals radii. NOE distances for most of the non-exchangeable protons were obtained from integration of crosspeak volumes obtained from the NOESY spectrum acquired with a 250 ms mixing time, using the average pyrimidine H5–H6 crosspeak intensity as a standard reference of 2.5 Å. Upper bounds were set at 20% of the NOE distance and lower bounds equal to the sum of the van der Waals radii. Additional non-exchangeable NOE distance restraints were obtained from the 3D NOESY-HMQC spectrum, which were classified semi-quantitatively as described above. The glycosidic (χ) angles were weakly restrained in the structure calculations to allow the entire *anti* range of –120 ± 90° for all nucleotides except G14, which the NOE data indicated is *syn* and which was restrained to encompass the entire *syn* range of 60 ± 90°. Dihedral angle restraints for the ribose sugar puckers (δ) were obtained from analysis of DQF COSY spectra; residues with small or absent H1′–H2′ crosspeaks were restrained to the N-type range (85 ± 30°), and those with intermediate coupling constants of 4–7 Hz were left unrestrained. The backbone dihedral angles α, β, δ, ε and ζ were unrestrained for all of the residues in the internal loop (5–7, 18, 19). A-form backbone dihedral angles (±10°) were included for the Watson–Crick stem regions (1–4, 9–10, 15–16 and 20–23), since the NOE data indicated that the Watson–Crick stems are indeed essentially A-form helices.

Structure calculations

All structure calculations were done with X-PLOR version 3.1 (Brünger, 1992). The structure calculation protocol used here is essentially the same as described by Dieckmann *et al.* (1996), with minor modifications. Hydrogen bonds were included for the six Watson–Crick base pairs in the stems, all of which displayed slowly exchanging imino protons as well as NOEs indicating the presence of these base pairs. Two hydrogen bonds were also included for the G8·U17 wobble pair, since the NOE data and an initial set of calculations without hydrogen bonds indicated that the G8·U17 forms a wobble base pair. The UUCG tetraloop was left largely unrestrained, since all NOE, chemical shift and coupling constant data indicated that the UUCG tetraloop assumes a structure identical to that which has been previously described (Allain and Varani, 1995b). In a typical calculation, 100 starting structures were generated using distance geometry full structure embedding. The structures were then subjected to a simulated annealing protocol of 15 ps at 2000 K, followed by 22 ps of cooling to 100 K with a time step of 0.7 fs, and 200 steps of energy minimization using the Powell algorithm. A square-well NOE potential was used for all calculations, with an initial scale factor of 50. A second refinement round of molecular dynamics and simulated annealing was then performed at 2000 K with cooling to 100 K in 40 ps with a 0.7 fs time step, followed by 200 steps of energy minimization. In this step the NOE and dihedral scale factors were increased to 100. At this point the structures were sorted by their NOE energy values and the convergence of the structures was examined graphically. Approximately 25% of the structures converged to low overall energies with no significant violations. The 20 lowest energy structures were subjected to a final molecular dynamics step, in which the full van der Waals term was introduced (16 ps at 300 K with a 0.2 fs time step followed by 200 steps of energy minimization). Structures were visualized and evaluated using the software packages MolMol (Koradi *et al.*, 1996) and Insight II (Biosym). Hydrogen bonds were analyzed with Insight II software, using criteria in which the angle between proton donor and acceptor must be greater than 120° and the distance between heteroatoms less than 3.5 Å.

Modeling calculations

Coordinates for the GAAA tetraloop were obtained from the Brookhaven Protein Data Bank accession number 1HMH (Pley *et al.*, 1994b). The tetraloop was docked by hand into the NMR structure using Insight II (Biosym). A distance constraint file was created consisting of the heteroatom distances for the hydrogen bonds between the tetraloop and the tandem C·G base pairs (Pley *et al.*, 1994b). The model was created with X-PLOR using 50 steps of rigid body minimization followed by 5000 steps of free energy minimization using the Powell algorithm.

Coordinates for the 20 lowest energy structures have been deposited with the Brookhaven Protein Data Bank, entry 1tlr.

Acknowledgements

References

Aboul-ela,F., Karn,J. and Varani,G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by TAT protein. *J. Mol. Biol.*, **253**, 313–332.

Allain,F.H.-T. and Varani,G. (1995a) Divalent metal ion binding to a conserved wobble pair defining the upstream site of cleavage of group I self-splicing introns. *Nucleic Acids Res.*, **23**, 341–350.

Allain,F.H.-T. and Varani,G. (1995b) Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.*, **250**, 333–353.

Allain,F.H.-T., Gubser,C.C., Howe,P.W.A., Nagai,K., Neuhaus,D. and Varani,G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature*, **380**, 646–650.

Batey,R.T., Inada,M., Kujawinski,E., Puglisi,J.D. and Williamson,J.R. (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.*, **20**, 4515–4523.

Battiste,J.L., Mao,H., Rao,N.S., Tan,R., Muhandiram,D.R., Kay,L.E., Frankel,A.D. and Williamson,J.R. (1996) α helix-RNA major groove recognition in an HIV-1 Rev peptide–RRE RNA complex. *Science*, **273**, 1547–1551.

Bax,A. and Davis,D.G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Resonance*, **65**, 355–360.

Bax,A., Clore,G.M., Driscoll,P.C., Gronenborn,A.M., Ikura,M. and Kay,L.E. (1990) Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of 13C-labeled proteins. *J. Magn. Resonance*, **87**, 620–627.

Brünger,A.T. (1992) *X-PLOR (Version 3.1) Manual.* Yale University Press, New Haven CT.

Burke,J.M., Belfort,M., Cech,T.R., Davies,R.W., Schweyen,R.J., Shub,D.A., Szostak,J.W. and Tabak,H.F. (1987) Structural conventions for group I introns. *Nucleic Acids Res.*, **15**, 7217–7221.

Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996a) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.

Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996b) RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.

Cheong,C., Varani,G. and Tinoco,I.,Jr (1990) Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature*, **346**, 680–682.

Chou,S.-H., Zhu,L. and Reid,B.R. (1997) Sheared purine·purine pairing in biology. *J. Mol. Biol.*, **267**, 1055–1067.

Clore,G.M., Bax,A., Driscoll,P.C., Wingfield,P.T. and Gronenborn,A.M. (1990) Assignment of the side-chain ^{1}H and ^{13}C resonances of interleukin-1β using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy. *Biochemistry*, **29**, 8172–8184.

Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276–1285.

Costa,M. and Michel,F. (1997) Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J.*, **16**, 3289–3302.

Dieckmann,T. and Feigon,J. (1994) Heteronuclear techniques in NMR studies of RNA and DNA. *Curr. Opin. Struct. Biol.*, **4**, 745–749.

Dieckmann,T. and Feigon,J. (1997) Assignment methodology for larger oligonucleotides: application to an ATP-binding RNA aptamer. *J. Biomol. NMR*, **9**, 259–272.

Dieckmann,T., Suzuki,E., Nakamura,G.K. and Feigon,J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, **2**, 628–640.

Jaeger,L., Michel,F. and Westhof,E. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.*, **236**, 1271–1276.

Jiang,F., Kumar,R.A., Jones,R.A. and Patel,D.J. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183–186.

Kadkhodaei,M., Hwang,T.L., Tang,J. and Shaka,A.J. (1993) A simple windowless mixing sequence to suppress cross relaxation in TOCSY experiments. *J. Magn. Resonance Ser. A*, **105**, 104–107.

Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL – A program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51+.

Legault,P. and Pardi,A. (1994) *In situ* probing of adenine protonation in RNA by C-13 NMR. *J. Am. Chem. Soc.*, **116**, 8390–8391.

Legault,P., Farmer,B.T.,II, Mueller,L. and Pardi,A. (1994) Through-bond correlation of adenine protons in a ^{13}C-labeled ribozyme. *J. Am. Chem. Soc.*, **116**, 2203–2204.

Macura,S., Huang,Y., Suter,D. and Ernst,R.R. (1980) Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J. Magn. Resonance*, **43**, 259–281.

Marino,J.P., Prestegard,J.H. and Crothers,D.M. (1994) Correlation of adenine H2/H8 resonances in uniformly ^{13}C labeled RNAs by 2d HCCH-TOCSY: a new tool for ^{1}H assignment. *J. Am. Chem. Soc.*, **116**, 2205–2206.

Marion,D., Ikura,M., Tschudin,R. and Bax,A. (1989) Rapid recording of 2d NMR spectra without phase cycling – application to the study of hydrogen exchange in proteins. *J. Magn. Reson.*, **85**, 393–399.

Maskos,K., Gunn,B.M., LeBlanc,D.A. and Morden,K.M. (1993) NMR study of G·A and A·A pairing in d(GCGAATAAGCG)$_2$. *Biochemistry*, **32**, 3583–3595.

Michel,F. and Westhof,E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.

Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) Oligoribonucleotide synthesis using T7-RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.

Murphy,F.L. and Cech,T.R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.*, **236**, 49–63.

Nikonowicz,E.P. and Pardi,A. (1993) An efficient procedure for assignment of the proton, carbon and nitrogen resonances in ^{13}C/^{15}N labeled nucleic acids. *J. Mol. Biol.*, **232**, 1141–1156.

Nikonowicz,E.P., Sirr,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) Preparation of C-13 and N-15 labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res.*, **20**, 4507–4513.

Pardi,A. (1995) Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. *Methods Enzymol.*, **261**, 350–380.

Peterson,R.D. and Feigon,J. (1996) Structural change in Rev Responsive Element RNA of HIV-1 on binding Rev peptide. *J. Mol. Biol.*, **264**, 863–877.

Peterson,R.D., Bartel,D.P., Szostak,J.W., Horvath,S.J. and Feigon,J. (1994) 1H NMR studies of the high-affinity Rev binding site of the Rev Responsive Element of HIV-1 mRNA: base pairing in the core binding element. *Biochemistry*, **33**, 5357–5366.

Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994a) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, **372**, 111–113.

Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994b) Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68–74.

Puglisi,J.D. and Wyatt,J.R. (1995) Biochemical and NMR studies of RNA conformation with an emphasis on RNA pseudoknots. *Methods Enzymol.*, **261**, 323–350.

Rance,M., Sørensen,M., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wüthrich,K. (1983) Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum altering. *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Robertus,J.D., Ladner,J.E., Finch,J.T., Rhodes,D., Brown,R.S., Clark,B.F.C. and Klug,A. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, **250**, 546–551.

Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitschler,A., Podjarny,A., Rees,B., Thierry,J.C. and Moras,D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA^{Asp}. *Science*, **252**, 1682–1689.

Saenger,W. (1984) *Principles of Nucleic Acid Structure.* Springer-Verlag, New York.

Santoro,J. and King,G. (1992) A constant-time 2d overbodengausen experiment for inverse correlation of isotopically enriched species. *J. Magn. Resonance*, **97**, 202–207.

Scott,W.G., Finch,J.T. and Klug,A. (1995) The crystal structure of an all-RNA hammerhead ribozyme – a proposed mechanism for RNA catalytic cleavage. *Cell*, **81**, 991–1002.

Serra,M.J. and Turner,D.H. (1995) Predicting thermodynamic properties of RNA. *Methods Enzymol.*, **259**, 242–261.

Shaka,A.J., Barker,P. and Freeman,R. (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *J. Magn. Resonance*, **64**, 547–552.

Sklenář,V., Miyashiro,H., Zon,G., Miles,H.T. and Bax,A. (1986) Assignment of the ^{31}P resonances in oligonucleotides by two-dimensional NMR spectroscopy. *FEBS Lett.*, **208**, 94–98.

Sklenář,V., Tschudin,R. and Bax,A. (1987) Water suppression using a combination of hard and soft pulses. *J. Magn. Resonance*, **75**, 352–357.

Sklenář,V., Peterson,R.D., Rejante,M.R., Wang,E. and Feigon,J. (1993) Two-dimensional triple-resonance HCNCH experiment for direct correlation of ribose H1′ and Base H8, H6 protons in ^{13}C, ^{15}N-labeled RNA oligonucleotides. *J. Am. Chem. Soc.*, **115**, 12181–12182.

Sklenář,V., Peterson,R.D., Rejante,M.R. and Feigon,J. (1994) Correlation of nucleotide base and sugar protons in ^{15}N labeled HIV RNA oligonucleotide by ^{1}H-^{15}N HSQC experiments. *J. Biomol. NMR*, **4**, 117–122.

Szewczak,A.A., Moore,P.B., Chan,Y.-L. and Wool,I.G. (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc. Natl Acad. Sci. USA*, **90**, 9581–9585.

Tanner,M. (1997) Bacterial group I ribozymes and their tertiary interactions. PhD Thesis, University of Colorado.

Tanner,M.A. and Cech,T.R. (1995) An important RNA tertiary interaction of group I and group II introns is implicated in gram-positive RNase P RNAs. *RNA*, **1**, 349–350.

Varani,G., Aboul-ela,F. and Allain,F.H.-T. (1996) NMR investigation of RNA structure. *Prog. NMR Spectrosc.*, **29**, 51–127.

Wimberly,B., Varani,G. and Tinoco,I.,Jr (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry*, **32**, 1078–1087.

Zimmermann,G.R., Jenison,R.D., Wick,C.L., Simorre,J.-P. and Pardi,A. (1997) Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nature Struct. Biol.*, **4**, 644–649.